

DOCUMENT RESUME

ED 150 265

UD 017 993

AUTHOR Thomas, Thomas C.; Muller, Jurgen
 TITLE The Efficiency of Educational Expenditures for Compensatory Education; An Interpretation of Current Evidence.
 INSTITUTION Stanford Research Inst., Menlo Park, Calif. Educational Policy Research Center.
 SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
 REPORT NO EPRC-6747-15
 PUB DATE May 72
 CONTRACT OEC-1-071013-4274
 NOTE 49p.
 EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
 DESCRIPTORS *Compensatory Education Programs; Data Analysis; Data Collection; *Educational Policy; *Expenditures; Federal Programs; *Measurement Goals; *Measurement Techniques; *Models; Policy Formation; Research Problems

ABSTRACT

Policy implications stemming from empirical analyses that evaluate national federal compensatory programs are clouded because negative findings could be due to either the nature of the educational process or the inadequacy of the data collection and analysis technique. This paper constructs a framework within which the separate effects of each are assessed. Two general sources of direct evidence are considered. The primary source is the result of multivariate analyses of national and state educational data such as found in Project Talent or the Equal Opportunity Survey. The secondary source is the results of analyses of national and state data on ESEA Title I project evaluations. It is found that standardized tests are useful for some purposes such as overall student comparisons, but their high interaction with student background factors blunt their usefulness in determining the most effective alternative school inputs, policies, and practices. Given the present output measures, it is concluded that variations in educational expenditures do in fact have little independent effect upon educational output. The main problems lie in the proper understanding of the educational process, the usefulness of the output measure, and the soundness of analysis techniques. A strong interdisciplinary attempt should be made to correct these problems. (Author/AM)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

ED150265

SRI Project 6747

May 1972

Educational Policy Research Center

THOMAS C. THOMAS

JÜRGEN MÜLLER

Research Note 15

EPRC 6747-15

**THE EFFICIENCY OF EDUCATIONAL
EXPENDITURES FOR COMPENSATORY
EDUCATION—AN INTERPRETATION
OF CURRENT EVIDENCE**

Prepared for:

NATIONAL CENTER FOR EDUCATIONAL
RESEARCH AND DEVELOPMENT
U.S. OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

CONTRACT OEC-1-7-071013-4274

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

UDO 17988

POLICY RESEARCH REPORT

A Policy Research Report is an official document of the Educational Policy Research Center. It presents results of work directed toward specific research objectives. The report is a comprehensive treatment of the objectives, scope, methodology, data, analyses, and conclusions, and presents the background, practical significance, and technical information required for a complete and full understanding of the research activity. The report is designed to be directly useful to educational policy makers.

RESEARCH MEMORANDUM

A Research Memorandum is a working paper that presents the results of work in progress. The purpose of the Research Memorandum is to invite comment on research in progress. It is a comprehensive treatment of a single research area or of a facet of a research area within a larger field of study. The Memorandum presents the background, objectives, scope, summary, and conclusions, as well as method and approach, in a condensed form. Since it presents views and conclusions drawn during the progress of research activity, it may be expanded or modified in the light of further research.

RESEARCH NOTE

A Research Note is a working paper that presents the results of study related to a single phase or factor of a research problem. It also may present preliminary exploration of an educational policy issue or an interim report which may later appear as a larger study. The purpose of the Research Note is to instigate discussion and criticism. It presents the concepts, findings, and/or conclusions of the author. It may be altered, expanded, or withdrawn at any time.

CONTENTS

EXECUTIVE SUMMARY	iii
INTRODUCTION	1
THE MODEL	4
The Societal-Educational Process	5
Educational Outcomes	6
Educational Process and the Lack of Learning Theory	10
THE DATA	16
THE ANALYSIS METHODOLOGY	24
RESULTS OF REVIEWED STUDIES	27
Other Reviews	27
Specific Studies	29
Effect of Expenditures	30
School Inputs	33
RECOMMENDATIONS	36
REFERENCES AND OTHER SELECTED STUDIES	39

EXECUTIVE SUMMARY

Nearly all empirical analyses of survey data on educational inputs and outputs such as Project Talent, the Equal Educational Opportunity Survey, and National ESEA Title I evaluations show little or no relationship between variations in educational expenditures and educational outputs. Nevertheless, the policy implications are clouded because these findings could arise in either the nature of the educational process or the inadequacy of the data collection and analysis techniques. This paper constructs a framework within which we assess the separate effects of each.

Our analysis of the educational process leads us to expect a low effectiveness of educational expenditures in the improvement of the most widely used measure of output, students' standardized achievement test scores. These test scores reflect the interactive effect of both school and nonschool influences. The independent effect of variations in school inputs is very low.

Standardized tests are useful for some purposes such as overall student comparisons, but their high interaction with student background factors blunt their usefulness in determining the most effective alternative school inputs, policies, and practices. Moreover, if they were to become educational goals rather than merely correlates of the goals, as could happen with the increased emphasis on accountability, they would be vulnerable to invalidation through "teaching to the tests." A next generation of criterion-referenced tests, which will overcome some of their present limitations, is required.

We strongly recommend that tests, measures, and indicators that are more sensitive to the effects of "good" formal education be created to provide a more useful basis for policy decisions about resource allocations within the public school system.

Given the present output measures, we must conclude that variations in educational expenditure do in fact have little independent effect upon educational output. The main problem lies in the nature of the educational process and the output measure, not the analysis techniques.

Nevertheless, serious problems exist with the analysis technique. Even though the effectiveness of educational expenditure may be low on the given output measure, it is still possible and important to estimate stable, statistically significant relationships to provide policy guidance. This has not been done.

No adequate models of the educational process exist from which to construct the relationships to be estimated. Research studies provide little information on crucial relationships at the level of the teacher and child. Perhaps even more damaging is the lack of information on the factors that influence the way different educational institutions actually allocate their effort in the face of multiple goals and changing priorities. Variations in allocative decisions among schools can effectively hide stable input-output relationship in an analysis focused on a single outcome such as reading achievement test score. Furthermore, without adequate models, it is difficult to identify the variables that should be measured and to pinpoint the biases that can be expected in the estimates due to the practical compromises that must be made in data collection. Every analysis of education contains at least an implicit model of the educational process, but the vague diffused nature of nearly all these "models" creates serious problems in the selection of variables and the assessment of biases in the available data.

The lack of well structured models plus the variety of partial variables--e.g., the presence of a science laboratory is taken as a measure of the overall quality of facilities--lead to the use of mechanical techniques to select variables. Some techniques, such as factor analysis, can be useful while others, such as stepwise regression or selective reporting, can invalidate the analysis for policy purposes. A majority of the studies reviewed fall prey either to the use of stepwise regression or to selective reporting.

A strong interdisciplinary attempt should be made to develop the best conceptual models and methodology for describing the educational process, based upon what is already known. The work to date has generally been oriented to one discipline, methodologically unsophisticated (and often simply invalid) and well below what could be done even given the current lacks in the supporting findings of education research.

Such a study would probably indicate the need for and potential benefits of new, specifically focused data collection efforts. However, it would also provide a referent against which to evaluate the biases expected in the use of the available data sources. We strongly recommend, whether new or existing data sources are used, that the potentialities of factor analysis or clustering technique be used to create useful macro-variables to reduce problems of multicollinearity and obviate the tendency to use stepwise regression or its variations.

It is not expected that the improved analysis techniques will result in finding a dramatically larger role for school inputs in the overall educational outcomes for the disadvantaged. However, school inputs are an influence and these inputs in total continue to cost multi-billions of dollars. Consequently, even small increases in our knowledge of how to use educational expenditures more efficiently could pay large dividends.

INTRODUCTION

Under the educational and societal conditions existing in the United States during the 1960s, the estimated potential of increased expenditures to make significant changes in measured outcomes of education for the disadvantaged was not practically significant. Nor was the capability of educational expenditures to improve education for the majority significant, although it was somewhat greater than for the disadvantaged. These two findings are frequently interpreted to mean that no educational benefit can be obtained from providing public schools with greater financial resources, a conclusion that is probably untrue and certainly needs to be as heavily qualified as the opening assertion; i.e., in the 1960s, estimated capability, measured outcomes, and practically significant.

The general acceptance by most researchers of the measured outcomes per se, in contrast to the disputes over the policy implications which can be drawn from them, has led us to concentrate on constructing a framework in which the implications and limitations of the empirical findings can be assessed. Specific findings will be used as illustrations throughout this paper and they will be considered as a whole at the end of the paper.

Figure 1 portrays the four main elements of an analysis: the model of the educational process, the data drawn from the process, the analysis methodology, and the results or estimated relationships.

- The model is the statement of how the educational process works. It is an abstract set of relationships which, when stated in a compact form, attempts to represent the essential elements of the process. To the extent that the statement of the model

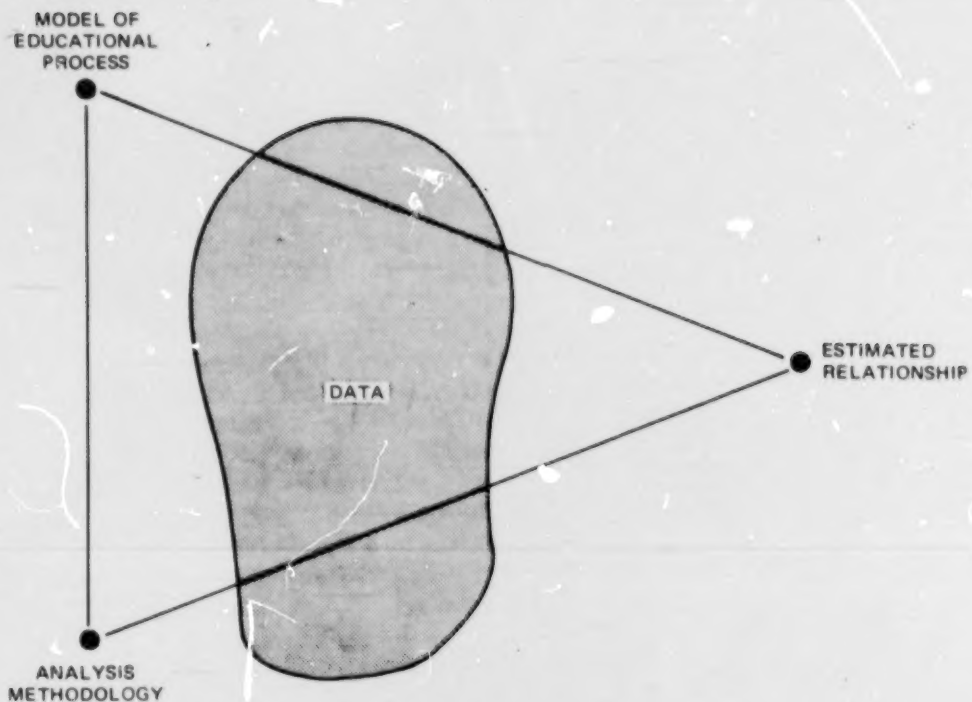


FIGURE 1

incisively portrays the crucial relationships, and only to that extent, can one have confidence that the results have predictive validity, i.e., that the results can be used to assess the effect of future policy changes on the variables included in the model.

The data form the link between the model and the real, multi-faceted, complex educational process under study. Any discrepancy between the data called for by the model and the actual data which are available will tend to cloud and obscure the true nature of the relationships that are estimated.

- The analysis methodology specifies the manner in which the data should be used to estimate the parameters of the model and thereby describe a specific educational process. At the current state of the art, the range of available analysis methodologies is limited and often the model is misspecified so that a particular estimation technique can be used at the cost of some increased estimation error.
- The estimated relationship should be viewed as the product of the first three elements. The implications of the results can be properly assessed only in terms of characteristics of the model, data, and methodology.

Of particular concern in this paper is an assessment of the extent to which the findings--changes in expenditures have had an insignificant effect upon educational outcomes--are due to the operational characteristics of the educational process and the extent to which they are due to limitations in the model, the data, or the methodology.

Two general sources of direct evidence are considered. The primary source is the result of multivariate analyses of national and state educational data such as found in Project Talent or the Equal Opportunity Survey (also known as the Coleman Report). The secondary source is the results of analyses of national and state data on ESEA Title I project evaluations. A variety of other sources are used when appropriate to provide supporting evidence.

THE MODEL

Educational models are of many different types, reflecting the different perspectives of the model builder. Thus, there are anthropological, sociological, psychological, and economic as well as eclectic models of education.

The economic models focus upon questions of the efficient allocation and utilization of resources and form the basis for most of the literature that was reviewed for this paper. Most economic models are input-output models. The objective of such analysis is to estimate the underlying relationships in the educational process between the inputs and outputs. Such an estimated relationship between educational expenditures and educational outcomes provides the basis for evaluating the efficiency of educational expenditures on education.

In economic terms, the educational process is seen as a production function by which inputs, e.g., unlettered students plus teachers and materials, are transformed into outputs, e.g., students who become enlightened and socially responsible citizens. As the goals of education change or the cost and quality of the different inputs change, the information obtainable from studying the production function indicates how the inputs should be changed to reach the new goals. The production function itself changes only when the underlying process of education changes.

Indications of change are difficult to capture, however, since the changes are often caused by political and social pressures, which are hard to quantify. This difficulty tends to reduce the usefulness of this approach, especially for policy recommendations for the disadvantaged.

The Societal-Educational Process

The educational production function is usually specified as some variation of the following general form:

$$Y_{it} = J(F_{it}, S_{it}, C_{it}, I_{it})$$

where Y_{it} is the vector of cumulative outcomes for the education of the i th student at time t .

F_{it} represents a vector of the student's family background characteristics at time t , whose effects may of course be cumulative. The specific elements of the vector include measurements of socioeconomic status (e.g., the parents' education, occupation, income, or racial composition) and specific aspects of the home life (e.g., the presence of an encyclopedia, quantity of books, sex of head of household or number of siblings).

S_{it} represents the vector of school resources going to the student at time t . Some of the most important school inputs are thought to be teachers' characteristics (e.g., experience, verbal ability, degree level, starting salary, or attitudes) and school resources (e.g., library volumes per student, age of building, size of school, or lab facilities).

C_{it} represents the vector of community endowments and other external factors such as student body racial composition, urban or rural environment, or occupations of peers' parents.

I_{it} represents the vector of initial endowments of the i th student at time t . At one time I.Q. test scores would have been used here without question, but there is no measure universally accepted at present.

It is assumed that the function is well behaved and twice differentiable, i.e., the Inada conditions hold and inputs may be chosen so that all the marginal conditions hold.

Within this general statement of the educational production function, the task is to select the precise variables and the form of the functional relationship so that the parameters of the relationship can be estimated. In this section, the specification of the outcome measures and the form of the educational process will be considered. Later, the section on data will treat the input measures.

Educational Outcomes

The goals of education are multiple, with priorities that change with the political and social environment and the individual's attitude. Of the specific subset of outcomes to be included in the model, two characteristics are particularly important: uniqueness and allocation of effort. Uniqueness refers to outcomes for which the public school system or certain elements in it are solely responsible. For example, the development of algebraic skills is uniquely attributable to the public school, while the development of basic verbal communication skills is shared with the family, peer group, and the mass media.

Allocation of effort refers to the percentage of the school's resources that are devoted to the outcome. For instance, only a limited number of students take algebra and their instruction in it is usually confined to a few years in junior and senior high school. The overall allocation of effort to algebra is therefore low. On the other hand, for basic communication skills, the overall effort is high and well distributed over the school's entire program, because communication is central to all other learning. In general one finds that those outcomes that have the highest allocation of effort across all students are the least unique to

the school and tend to be shared with other elements of the society. Since these outcomes are also the ones on which evaluations focus, the problem of attribution becomes a very difficult one indeed.

The problem of estimating the school's independent effect increases still further when we consider the difference between potential and actual outcomes. If factors that reduce the potential outcomes are not properly specified, inferences become meaningless. When we keep in mind how large the discrepancies are between potential and actual outcome in the case of the disadvantaged, the difficulty of estimating the "true" relationships becomes apparent.

The actual choice of specific output measures in the studies reviewed was not made on the grounds of uniqueness, allocation of effort, or implementing process, but rather for the statistical properties of reliability and validity and, in the end, mainly for ease of collection across large samples of students. The measures are mainly the results of standardized cognitive tests, especially verbal ability, reading achievement, and mathematics achievement.

It is important to realize that although some of these tests are labeled achievement tests, they are more aptly thought of as tests of general ability. The schools do not specifically teach the material in these tests; indeed, given the basis on which the tests are constructed, "teaching to the tests" is specifically enjoined. The schools' inputs may be somewhat more unique in the area of mathematics than in the area of verbal ability, but overall what these tests measure is not uniquely learned in the schools.

This situation is graphically illustrated by Mayeske's analysis of the Equal Educational Opportunity Survey data.¹* While the precise results change with the grade level, he found socioeconomic status (SES)

* The references are listed at the end of this paper.

variables had roughly twice the unique influence of the school inputs and that the interaction of SES and school inputs had more than 13 times the influence of the unique school input. Thus, from the point of view of uniqueness, one would expect that the independent effect of increased school expenditures would be very low when compared with situations in which SES or SES plus school inputs change.

This raises the very important point that educational problems are not solely school problems. Most attempts to improve school outcomes focus on school input alone, but strategies that concentrate on outside factors of, for instance, SES and health² may produce better results.

Some insight into this lack of a unique school effect can be gained by looking at the differences between standardized and criterion-referenced tests. Criterion-referenced tests are specifically constructed to test the student's performance on the learning objectives of a particular instructional block. Thus a criterion-referenced test in American history would be constructed to test the student on his knowledge of American history as taught in that course. This material would tend to be unique to school. One would consequently expect that expenditures allocated to increasing the students' outcomes on the American history criterion-referenced test would have a high potential for effectiveness.

Unfortunately, there are two major problems with criterion-referenced tests. The first is the obvious difficulty of classifying the various schools' curriculums and creating a separate criterion-referenced test for each--and of comparing and weighting these scores. The second and more serious problem from a modeling perspective is that the criteria-referenced tests do not predict future school success very well, not even outcomes on other criterion-referenced tests. Standardized tests are much better at predicting future school performance, not surprising when one remembers that the "developer's basis for improving (standardized) tests has been to work toward better predictions of later performance rather than better

measurement of present performance. Assessment of what a student is now capable of doing is not the purpose of most standardized tests."³

The conclusion appears to be that on the predominant outcome measure, the results of standardized tests, the school has little unique influence. The interlocking and redundant inputs of school and society would perhaps not be very surprising if they applied to the overall life opportunities of the student, but it is more surprising that the schools' independent effect upon school performance is so low. This finding is further explored later.

The next question is how important these cognitive test outcomes are in terms of the allocation of funds within the public schools. This is a difficult question to answer precisely. The tests range widely over areas of general cognitive knowledge, but are limited to a timed, paper and pencil motif. Moreover, it is clear that the public schools have many other goals which compete with specific areas of cognitive learning for available funds. Probably the most interesting data available for our purpose relate to the allocation of funds under ESEA Title I. These funds represent a large and visible increment of money, specifically for increasing the cognitive skills of the disadvantaged, money which had to be separately accounted for.

The first thing that happened to much of the Title I money was that it encountered multipocketed budgeting or symbolic allocation. This is the practice by which funds appeared to go where they were intended, but in effect were used to release, for use in other areas, resources that would have otherwise gone to the disadvantaged. As a consequence, the net increase in funds was much smaller than the nominal increase. No precise figures exist for the quantitative impact of this practice, which was strongly discouraged by USOE, but a study by the Syracuse EPRC⁴ found

it prevalent in most large districts and in those smaller districts with the sophistication and budget complexity necessary to hide this maneuver.*

Of that money allocated to the disadvantaged, Wargo and coworkers found in an analysis⁵ of 15 states which provide data on expenditures divided into instructional or support program categories, that from a low of 10 percent to a high of 50 percent of the funds were expended on support areas, e.g., clothing, food or health. The average in FY 70 was 26.3 percent. Moreover, in most of these support areas, the number of students serviced was greater than the number in need and most of the students were not disadvantaged. However, the vast majority of the funds spent in the instructional areas were targeted on the disadvantaged.

While the support funds, when used for the disadvantaged, could be supposed to have long run effects on cognitive achievement, the overall impression is that appreciable funds are diverted away from cognitive achievement. When this finding is considered with that on uniqueness, the schools' independent potential for raising cognitive achievement scores appears to be very low. (And they have only limited freedom in the allocation process itself, given state regulations.)

Educational Process and the Lack of Learning Theory

The question that is raised here is not whether the schools know how to educate children, for indeed children do become educated in schools. The question is rather whether the state of the art is such that the schools can be expected to direct efficiently the utilization of both existing and additional resources. In the educational literature, one finds that exemplary programs are developed in the schools, but it is

* Multipocketing would bias the results of an experimental-control group analysis of Title I if the multipocketed money was spent on the control group.

often hard to tell what indeed is exemplary and exportable about them, questions that research findings should be able to answer.

Unfortunately, if one turns to the area which has received both the most emphasis in research and is the most relevant to cognitive outcome measures, namely reading, the findings are ambiguous at best. The Educational Testing Service, in its work for USOE's Targeted Program of Research and Development in Reading, studied over a thousand documents hand-picked for their potential significance, and reported that, "unfortunately, the conclusion is that the present literature on methods of reading instruction cannot be used, except in special circumstances, to justify involving millions of children..."⁶ Regarding the specific instructional materials which are publicly marketed, the report found that less than ten percent had been field tested and only about one percent had been subjected to learner-verification tests. While the report did identify one or two hopeful signs for the future, to date there is very little evidence that the schools have any source of tested knowledge on how to use additional funds to improve their reading programs.

A much wider ranging study of what is known about how to teach effectively was done by Averch, Carroll, and Donaldson.⁷ They concluded with respect to the classroom studies they examined that:

- (1) The research on teaching approaches, teacher differences, class size, and the like shows no consistent effect on student achievement, as measured by standardized cognitive tests. Other measures are seldom used.
- (2) The work on student ability and grouping, or tracking, indicates that IQ makes a difference in achievement; but IQ is not directly susceptible to policy instruments and there is an interaction between IQ level, instructional techniques, and teacher character. In other words, high-IQ students will do better with

some kinds of teachers than with others, controlling for instructional techniques.

- (3) Work on instructional methods suggests no difference between methods; none currently appears better than conventional methods. That is, in terms of differences of achievement, conventional methods appear as effective as, say, teaching by television, although the latter enables exposure of a far greater number of students.

In sum, they were not able to identify the specific student relations involved in learning and education. "There are interactions between students and teachers, between students and methods, between teachers and methods, and (most complex of all) between students, teachers, and methods."

Studies of the education production function could help to identify a relationship between input and output and thus point out the nature of the underlying structure. But even without the data problems mentioned above, these studies do have shortcomings. Averch et al. concluded (p. xiii):

The complex three- and four-way interactions have not yet been studied carefully, but enough is known to suggest that they are one of the reasons that education research has difficulty delivering consistent results, particularly using the input/output approach. The latter assumes that resources are resources to students of all kinds, or to groups of students. It does not recognize that what is an educational resource to a particular student may not be one to another. If it did recognize this possibility, a difficulty would arise in fitting the statistical relations that are the sine qua non of the input/output approach.

The schools do not appear to know how to handle these interactions. More important, the extent to which schools even consider optimizing interactions at a practical level is not clear.

Still, production function studies do not require that we know the precise objective functions of schools. As long as we know that they are maximizing outcomes to meet some goals, they will be observed on a production frontier and the data should describe such behavior. However, these goals may not be taken into consideration by an outside observer, who may thus conclude that inputs used have little effect on outcome, simply because some outcomes are not being counted.

A case in point is the sorting role of schools. Schools in general appear quite satisfied to classify students according to how well they do at a set of relatively uniform intellectual and social tasks. Implicit support for the thesis that the sorting aspect of schooling is very important, and possibly dominant in the long run, comes from the standardized nature of the teaching methods (thereby making the student fit the mold), the greater predictive ability of the standardized tests compared with criterion-referenced tests (the schools don't teach the students how to succeed in the schools), and the low unique effect of the schools on the standardized results.

The schools, current rhetoric to the contrary, are infrequently individualized but rather sort out good and poor learners on how they fit the standard educational methods. This approach of the schools is particularly inefficient when applied to the disadvantaged who by definition are not fitted by their homes and peers for the standard tasks.

The main ramification of this view for the efficiency of educational expenditures is that the stronger the sorting process in school, and the weaker the school's attempts to provide compensatory aid to students in meeting the sorting criteria, the less effective will be increased education expenditures for changing the sorting experienced by the disadvantaged student. At the extreme, the money could go to improve the sorting process, not the learning process.

In summary, we would expect a low efficiency of educational expenditures on the improvement of students' cognitive achievement test scores. This may be due to low uniqueness (other, non-school factors need to be changed as well) and to the fact that much of these funds is spent to achieve other than the measured outcomes. These "other" outcomes may have an insignificant or even deleterious effect on the disadvantaged. Even when the money is specifically allocated for the cognitive achievement of the disadvantaged, there appears to be substantial leakage. Thus the overall potential for improvement is low.

However, inadequacies in current models could well hide some relationships that do exist. Even though the effectiveness of educational expenditures is judged to be low, this does not preclude the estimation of stable, statistically significant, relationships. The main problem in the models appears to be their inability to account for the variability in the operational-institutional goals different schools emphasize.

There is at present no model of the goal and resource allocation process which can generate estimates of the effect of variations in allocation of effort for each student in a school, or even for each school. Consequently, even if a stable relationship existed between each dollar of expenditures on cognitive achievement and test results (given the students' SES, peers, and initial intellectual endowments), the estimated coefficient of expenditures in a sample of schools devoting different percentages of their expenditures to cognitive goals would tend to have a large standard error and could well fail to pass a test of statistical significance.

Only a model general enough to incorporate the variability of goals could properly estimate the assumed stable relationship. The major problem is not one of achieving outcomes--because inputs are assumed to have effects--but one of handling the variety of different outcomes that are achieved.

Models of the allocative processes of education institutions faced with multiple goals and changing priorities should be developed. Such models are crucial to the estimation of input parameters for subsets of outcome goals. This research might have the added benefit of indicating to schools the advantages of longer range educational programming.

THE DATA

The central data problem rests in the absence of a body of educational theory and research findings from which to specify a detailed analytical model of the education process. Such a model would identify the variables that should be measured and, when practical considerations force compromises to be made in data collection, it would point to the biases that can be expected in the estimates. Every analysis of education contains at least an implicit model of the educational process, but the vague diffused nature of nearly all these "models" creates serious problems in the selection of variables and the assessment of biases in the available data.

Most discussion of data limitations in the studies reviewed for this paper focused on problems with specific elements of data and, indeed, there is hardly any limit to the number of such problems that can be uncovered. However, the most serious problem, one which poses a major threat to the predictive validity of even the limited findings in the studies reviewed, occurs in the selection of variables.

The general form of the production function model makes educational outcomes a function of the quality of the student's interaction with his family, peer group, teachers, and school facilities. The school interactions are partially controlled by the quality of the administrative, personnel, and operating procedures of the district. At this level of model specification, there are hundreds of relevant variables that have been collected, e.g., in the EEOS, and literally thousands that could be measured as representing the quality of the student's educational interactions and consequent outcomes. For many reasons, e.g., lack of degrees of freedom and multicollinearity, it is not desirable to use all

these variables in the analysis and somehow a subset must be selected.

At this point there are four general approaches:

- (1) Select the subset of variables based upon theory and judgment, and then estimate the coefficients and test the model.

This is the preferred approach and when well done ensures the methodological validity of the analysis. Some studies have attempted this approach, but the problem is that the theory is usually not worked out well enough to provide for a detailed selection of variables. Consequently, the analyst often turns to some mechanical technique to select the variables.

- (2) Perform a factor or principal components analysis to create aggregate variables.

In this approach, the analysis takes advantage of the fact that a number of the specific variables are indicators of essentially the same general factor and attempts to create a condensed measure for that factor. For example, in EEOS, the quality of school facilities is indicated by such variables as library volumes per student, percentage of makeshift classrooms, specialized rooms and facilities, science lab facilities, and separate classes for special students. In a factor analysis, each of these variables could be simultaneously analyzed to construct one or more principal components which capture the commonality of the variables. Ideally, a single overriding principal component would emerge that would be readily identifiable as an aggregate measure of the quality of the school's physical facilities. This single variable would then be used in the analyses in place of the all individual measures of school facilities.

This approach creates no methodological problems of validity in the estimates of educational outcomes because the input variables are constructed independently of the outcome variables. Unfortunately, only one study (Mayeske)² that was reviewed took this approach.

The main limitation of factor analysis is that the exact character of the constructed variable can at times be clouded and it may be difficult to interpret in policy terms the regression coefficients of the constructed variable on the outcome (dependent) variable.* However, this is a problem with the model structure in the first place; mechanical manipulations of data are complements, not substitutes, for theory.

(3) Use stepwise regression.

The essential feature of this approach is to see how well a specific variable "explains" the outcome variable in an early step of the analysis, so as to determine whether the variable is used in the final analysis. The (independent) input variables are regressed one by one and in combinations until the best overall fit is obtained. This approach violates the basic conditions for statistical significance and is void of any theoretical content. The end result is not an estimated model, but mechanical curve fitting. The results provide little basis on which to make statements even about the underlying structure of the data which were fitted and no basis on

* If the factor analysis tends to be unstable as a function of the number of factors identified, a more stable variable can often be developed by a cluster analysis of the same data.

which to predict the effect of major policy initiatives that create new circumstances.*

A large number of the existing studies must be criticized on the basis of using this technique.

Different studies have shown science lab facilities explaining reading scores, or dictionaries in the home explaining math scores. The author of such a study usually states the obvious, that the variable is only a proxy, but a proxy for what? With this approach, science labs could be a proxy for racial composition of the schools as well as facilities quality. Perhaps for "research," the stepwise approach can be justified as a quick and dirty alternative to factor analysis (though we think not), but it destroys any validity of that portion of the study for policy analyses.

(4) Run a number of regressions.

This approach specifies a number of different models and estimates each, and then tests their explanatory power and the significance of the coefficients. It is a multiple application of the first approach. However, a problem arises if subsequently only the few best estimators are selected for reporting; then, in effect, the approach becomes another form of the stepwise

* This is strictly true only for a single study. If other studies using different data and stepwise regression find the same variables to be significant in explaining outcomes, then there is some basis for assigning explanatory power to the variable. However, the coefficients will normally be different in the different studies and the question becomes: what is the true coefficient of the variable? Its true coefficient need not even lie between the estimated coefficients.

regression and is subject to the same criticisms.* Ideally, if there are doubts about the structure, different samples should be used--one on which to test the structure and another on which to estimate the coefficients of the previously chosen structure.

In summary, then, the benefit of using mechanical techniques to select variables as a substitute for theory is severely limited. Some techniques, such as factor analysis, can be useful while others, such as stepwise regression, can invalidate the analysis for policy purposes. A majority of the studies reviewed fell prey either to the use of stepwise regression or to selective reporting.

Turning to the quality of the individual data series, the variables of most immediate relevance are educational expenditures. The accounting systems at the school district level are "official," audited, and reasonably accurate within the classification system they have chosen, e.g., personnel, maintenance, or capital improvements. However, their accuracy and usefulness at the individual school level is abysmal. The PPBS thrust in education is directed towards the correction of this situation, but it certainly has not yet reached its goal. A few of the studies reviewed went into schools' accounting systems in detail in an attempt to pull out the specific data needed and verify their accuracy (e.g., Burkhead⁹) but most accepted the readily available data. It is difficult to estimate the overall impact of the inaccuracy in expenditure data, but the general effect would be to decrease the coefficient of the

* If all the estimators are published, this criticism is muted, but it is well to remember that on the average, one out of 20 estimators using random numbers for explanatory variables will be significant at the 5 percent confidence level (by definition of a confidence level). One study (Thomas),⁹ which used 18 different estimation equations and reported only one, can be criticized on these grounds.

expenditure variables (the effect of superimposing any unbiased error upon the true data series is to lower its estimated coefficient). For the disadvantaged, the general findings in the Title I comparability audits suggest that, within a district, the reported expenditures by school would tend to be biased low in the high SES schools and high in the low SES schools. This would tend to increase the significance of SES (as a correlate of school expenditures) and decrease the independent significance of expenditures.*

Within the school, the actual inputs for each individual student, e.g., teachers, books, classrooms, instructional programs, are seldom used (because they are generally not available), and the variables are overall school averages (an exception is Hanushek's study¹¹). This averaging out assumes that all students are affected equally. If this is not the case--and evidence suggests that differences exist--the estimated coefficients will be biased, depending on the average deviation.

Another problem arises from student mobility. Most urban disadvantaged students are highly mobile; many urban disadvantaged schools have a 100 percent turnover in a year's time and turnovers in excess of 50 percent are common. This poses great problems for creating appropriate treatment variables. If the study chooses to recognize the problem, it generally handles it by restricting the sample to students who have been in the school for the entire period under study. However, such a sample is clearly atypical. The home as well as the school situation could be expected to be less disrupted than normal and hence more effective.

* This applies to multivariate analyses, not Title I evaluations. The problem with Title I evaluations is that it is very difficult to track funds into the individual schools and obtain program costs accurate enough for any type of cost-effectiveness analysis.

This problem points to a more general one. Many of the inputs and outputs have a cumulative effect. Unless we can observe two points during the education process, it is very difficult to estimate any unique relationship between outcomes and inputs during a specific period. Only Hanushek¹⁰ used such a "value added" approach. In general, there exists very little data to perform such a longitudinal study.

Many other problems exist. The following are quoted (pp. 5-6) from Kiesling's study of multivariable analyses.¹¹

1. Teaching characteristics have been described by standard items in personnel folders--degree level, years of experience, salary, and the like. An adequate theory of instruction would require data concerning types of instructional techniques, time spent on them, order of presentation, etc. But even without such exact analysis, the teaching variables leave much to be desired. For example, data on basic teaching techniques would be useful. Information is also needed on the teacher's college major, grade point average, emotional stability, and intelligence.
2. In many instances a good variable has not been available for class size. Pupil-teacher ratio has often been the proxy used. Because staffing patterns vary, pupil-teacher ratios can be misleading as an indication of class size.
3. There have been virtually no variables available concerning teaching techniques used inside the class; all too often even such simple distinctions as group versus individual instruction cannot be made.
4. Cognitive test scores are not cardinal measures and are not very meaningful for "nonaverage" children.
5. Socioeconomic status (SES) variables have often been suspect, especially when used for small children. SES variables in the Equal Opportunity Survey were biased, perhaps considerably, because of a poor response rate.
6. There have been no good measures of management quality or of school physical inputs. The variables that have been available in these two areas have often been dollar

aggregates (expenditure in, or value of, plant and equipment, etc.) and the researcher has no way of knowing for certain how the dollars were spent.

7. Even variables for race have not been as good as many seem to think, since they normally do not give patterns of racial mixture in the classroom.

Lastly, the problem of restricted range should be mentioned. If an input variable is constant there is no way to estimate its effect upon outcomes (which is very different from saying that it has no effect). Other things being equal, the greater the range of variation in an input variable, the more precisely its effect, i.e., its coefficient, can be estimated. Restricted range appears to be a problem in some studies. A related but different problem is that, even if the coefficient is statistically significant, the range of the variable may not be large enough to allow variation from one extreme to the other to have a major impact upon the outcome.*

In summary, the lack of a well specified model of the educational process creates many measurement problems. Serious threats to validity of the model result from methods, primarily stepwise regression, which select the independent variables solely on the basis of correlations with the dependent variables. In addition, there are enormous data collection problems. These degrade the estimated significance of the expenditure and other school input variables.

* For example, teacher ability has been found to have a significant effect in analyses of EEOS data. However, as Michelson¹² has shown, the maximum change in this variable in favor of black students would only reduce the percentage of white students who rank above the black students average achievement level from 78.5 percent to 71.4 percent.

THE ANALYSIS METHODOLOGY

A number of the methodological questions have already been touched upon in the previous section. Prime among these is the use of stepwise regression. Another problem is multicollinearity. Many of the input and output variables in education move more or less together over a relatively narrow range, making the identification of unique contributions very difficult. Even though there are some techniques (such as first differences) to remove this problem, it complicates the model building and testing technique considerably.

Actually, most studies are based on rather simple additive models. These models may be misspecified. The coefficient of an expenditure term in such a model would, if taken seriously, state that expenditures have an effect upon output which is independent of all other considerations, e.g., a \$100 increase in per pupil expenditures would increase reading scores by one month per year independent of SES, teacher ability, or facilities. This is a most unlikely assumption, one that can be at least partially reduced by use of a log linear form of the equation. The assumption of constant elasticities may also be unwarranted but it would appear to be a little closer to the real relationships.*

In a situation in which the variables are interactive, the use of an equation system that can trace the effect of these interactions is

* For curve fitting over a narrow range of the data, there is often little empirical difference between the use of a linear or a log linear estimator, although most of the authors who report testing the log linear form appear to prefer it. Kiesling especially obtained good results using this form.¹³

indicated. A few attempts have been made to use such systems with some interesting results (Parti and Adelman,¹⁴ Levin¹⁵). On the basis of Levin's work, the conclusion is that the usual single equation estimators of cognitive test outcomes produce significantly biased results.*

Mention should also be made of the problems in using estimates from cross-sectional data for policy analyses. Cross-sectional data provide estimates of long run effects but can well hide the all important short run dynamics. The school districts at the 60th and the 80th percentile in expenditures may well have been at that level for 10, 15, or more years. Thus a comparison of the school district at the 60th percentile with the one at the 80th percentile will indicate the long run effects to the lower school district if its expenditures are shifted up. However, it does not indicate what the short run (2 to 5 years) effects will necessarily be, and these effects are of most interest to policy makers. The dynamics are better estimated from longitudinal data which show the effects as school districts change over time.

In summary, then, there are a number of methodological problems. None appear to be insoluble if it is possible to develop better models of the educational process. The fact that the various studies to date have had difficulty in showing that increased school resources have any significant effects appears consistent with our analysis. There are a number of measurement problems that invalidate some studies and tend to hide the effects of school inputs in all studies. Among these are:

* A single equation estimator assumes that the dependent variable is affected by the independent variables but does not affect them. When this is not the case, it can create appreciable biases in the estimates of the coefficients.

- (1) We know little of how to model the allocative process among various goals within school systems and hence will continue to have problems with measuring the overall effectiveness of increased resources. This difficulty is less important for categorical funds, but by no means insignificant.
- (2) The use of stepwise regression and variations of this technique results in generally useless probes for true aggregate variables (e.g., school facilities) and, in the process, destroys all the policy-predictive validity that the study might have. The result is the wide inconsistencies in the coefficients of variables that exist across studies, which casts doubt upon all of them. A principal components analysis (as used by Mayeske¹) can reduce the problem to some extent but it is not a substitute for a well specified model.
- (3) There are definite problems of multicollinearity. However, these problems have been conquered in the past when strong theoretical foundations were available to aid in specifying more sophisticated applied models. The work of Parti and Adelman¹⁴ is indicative of the approaches that can be taken when a strong theoretical foundation is available.
- (4) Cross-sectional data tend to hide the dynamics resulting from new policy recommendations. Longitudinal data are needed to study the dynamics.

In light of these operational and measurement considerations, we now look at some of the published results.

RESULTS OF REVIEWED STUDIES

Other Reviews

During our own analysis, we studied similar comparative analyses conducted by other groups. In particular, two Rand working papers--one by Kiesling¹¹ for ASPE of HEW, and the other by Averch, Carroll and Donaldson⁷ for the President's Commission on School Finance--are significant for this discussion. Also, the findings of the book edited by Mosteller and Moynihan,¹⁶ which is based upon a four-year analysis of the EOS by the Harvard Education Seminar, are generally consistent with the conclusion of the Rand papers and our own analyses, i.e., that little evidence currently exists to support the importance of variations in inputs of school resources in student outcomes.

One other book, by Guthrie, Kleindorfer, Levin and Stout¹⁷ reports on an apparently similar survey but arrives at the opposite conclusion; they "are impressed with the amount and consistency of evidence supporting the effectiveness of school services in influencing the academic performance of pupils." (p. 84). However, their selection of material and presentation of findings is clearly not an attempt to provide a balanced analysis of the state of the art. Rather, it is an attempt to support their central contention that the financing of public schools needs reform--a related but distinct question.

Returning then to the main thrust of the findings, the Averch, Carroll and Donaldson summary of input/output (production function) analyses is illustrative. They looked at 18 major research reports

2

comprising over 75 distinct studies, somewhat more than we were able to cover.* They state (pp. ix) that:

Looking at each study individually, [for internal validity] we find that:

- (1) Explanatory power is seldom impressive.
- (2) Background factors such as socioeconomic status are always important.
- (3) School resources are seldom important.

Looking across studies [to establish external validity]:

- (4) Background factors are consistently important--that is, they show up in study after study.
- (5) School resources are not consistently important--that is, a school resource claimed to be effective in one study does not often show up in another study, even when the resource is measured in roughly the same way.

In short, the input/output approach provides very little evidence that school resources, in general, have a powerful impact upon student outcomes, even neglecting the question of whether particular school resources are influential.

We want to emphasize that the results from the input/output approach must not be interpreted as indicating that school resources do not or cannot affect student outcomes. We observe that so far these studies have failed to show that school resources do affect student outcomes.

In addition, they also examined the evidence of effectiveness of ESEA Title I expenditures. The general conclusions are consistent with the findings of Wargo, et al., namely, that there is no persuasive

* Their report⁷ is particularly useful in that it contains an appendix with a summary analysis of each study.

evidence of effectiveness for the average ESEA Title I project. Effective projects do exist but they are the exception.

Specific Studies

Of the individual studies reviewed, only three--Levin,¹⁵ Michelson,¹² and Parti and Adelman¹⁴--used the simultaneous equation system approach. Levin's and Michelson's main interest appears to have been in testing the effect of the presumed misspecification in the single equation least squares model (Levin and Michelson used the same data and variables). Several modifications in the effects of particular variables between the simultaneous and single equation models were found by Levin and indicate the potential importance of the simultaneous approach, but it is still only a step in the right direction for correct model building. Levin notes that the results should be treated primarily as indicative.

The Parti and Adelman analysis was much more concerned with the specification of the model on a firm theoretical basis. Both their careful model specifications and the empirical results are excellent. However, the theoretical basis for the model--an outgrowth of the theory of consumer choice applied to the student--appears too limited to provide a firm basis for dealing with broad questions such as societal-educational interactions and the nature of institutional allocations.

These studies are described because we believe future efforts should go beyond the single equation approach and toward the simultaneous equation approach. The current results have little to say about the efficiency of educational expenditures per se. Levin and Michelson use library volumes per student as their only measure of quantity of school inputs, and Parti and Adelman "bury" class size in their causal chain.

Effect of Expenditures

Only two studies, Kiesling's study of large urban schools in New York State based upon their 1958-60 quality measurement data¹⁰ and Thomas's study of Project Talent data⁸ show major expenditure effect. Kiesling estimates, in a log linear form, a gain in average achievement per \$100 of expenditures per pupil which ranged from 1.4 to 2.6 months per year for large urban districts. The relationships for all urban districts combined and all other (non-urban) classifications separately were not predictable. Moreover, from his later work,¹¹ Kiesling concluded (pp. 30-31) that

per-pupil expenditure...was never highly related to pupil performance when socioeconomic differences were held constant. The exception was large urban school districts in the 1960 New York study. When all the Project Talent high schools are considered (Cohn)¹² the expenditure variable is strongly related to performance, but when the schools are stratified by type (urban, village, etc.) or region, the relationship disappears. This finding would imply some combination of poor estimating equations and wide variations in efficiency. Part of the explanation could also be that there was no way in which the estimated model could deal with existing differences in cost of educational resources.

Thomas's study⁸ used Project Talent data and found that a one standard deviation change in expenditures or \$144 per pupil (mean = \$334 per pupil) resulted in a 0.4 standard deviation change in student's achievement on the information test. However, this was for the one equation he published out of 18 equations he estimated, and his study is open to serious question on these and other grounds. Moreover, other studies using Project Talent data do not support his conclusions, e.g., Cohn.¹³

Studies using EEOS data do not show expenditures per student to be significant. Coleman, in his original (much criticized) study,¹⁴

found only 0.3 percent of the total variance in overall student outcome related to variations in expenditures (3 percent for Southern Negroes). These findings were recently reaffirmed by Smith in his reanalysis of the Coleman data.²⁰

A number of other studies looked at expenditures per pupil but often eliminated the expenditure variable in a stepwise regression procedure. Most of the studies used linear additive estimators. Thus, the methodology was generally poor and the results negative. Of even greater significance is the fact that when Kiesling stratified his data into five SES levels,²¹ he found that there was an increasingly negative effect of increased expenditure as SES decreased, i.e., for the lowest SES categories, the coefficient turned negative.

Using Project Talent data, Ribich²² also showed decreasing marginal returns from expenditures per student as SES decreased. For school inputs in general both Michelson² and Hanushek²³ found that the coefficients for low SES and minority groups, even if positive, were less significant than for majority students. Thus the low effectiveness of expenditures for all students is apparently compounded for low SES and minority groups. As disappointing as such a result may be, it is completely consistent with the evidence on the interactive effects of school inputs and student background, as well as with the thesis that sorting on the basis of student outcomes in standardized learning situations is a major function of the school.

The AIR study of the effectiveness of ESEA Title I programs⁵ provides corroborating evidence. When comparable data for reading achievement from the 1968 and 1969 nationwide surveys of Title I projects were combined, the overall result was that Title I did no better than the comparison group. This result is flawed by problems with (1) the comparison group, e.g., they were not randomly chosen and their initial

reading scores were appreciably higher than those of the treatment group, and (2) the small proportion--9 percent--of the total reported data that could be used, which poses a challenge to the representativeness of the data.* In another part of the analysis, AIR examined the Title I projects which were reported by five States as the ten most successful and the ten least successful. The ten least successful had a slightly higher average cost than the ten most successful. This amazing finding is possibly the result of multiple goals in the more expensive program as well as the lack of knowledge on how to teach the disadvantaged.

However, it is not clear that the situation is as bad as it might appear. Kiesling²⁴ developed a rough model of the most important elements in Title I projects and made a multivariate analysis of a sample of Title I projects in California on data which he collected and verified. He was able to show significant coefficients for measures of instructional intensity and instructional design. This study offers hope that the application of more sophisticated analysis tools to Title I projects may identify significant relationships for policymakers.

* In another analysis of five separate State reports, chosen as probably unbiased, approximate monthly gains in reading for a month in a program were reported. However, no control groups were reported and the increasing deficit on the pretest scores with grade level indicates that the disadvantaged students were not maintaining their relative position. This is possibly an artifact of pre- and post-testing which is not longitudinal and omits the summer (the testing is within the school year). It appears that disadvantaged students may suffer a 2- to 3-month decrement in achievement scores between May and September while other students hold relatively constant. This decrement is then reflected as an "easy" gain at the start of the year and produces an upward bias in the estimated improvement related to compensatory programs.

Unfortunately, Kiesling priced out these variables and his findings have been used by others to assert that Title I programs with funds concentrated at the level of \$300 per pupil are effective. As a general assertion, this appears unwarranted. Rather, the evidence indicates a complex of actions, of which funding is only one that might lead to successful Title I programs.* The AIR study⁵ also reviews lists of factors most often found in successful programs.

All above analyses should provide useful inputs for constructing better analytical estimates of the effectiveness of different Title I programs.

School Inputs

Most studies have chosen to use specific school inputs (e.g., library volumes per student) or expenditures in specific categories (e.g., expenditures on materials and supplies or on school administrators) in place of a total expenditure variable. Stepwise regression is commonly used in these instances and the significance of any one variable is erratic across studies. The exception to this rule is found in variables reflecting teacher inputs.

In some studies, teacher salary has been found to be an important variable but this would appear to be primarily a measure of which districts can compete to hire the best teachers. Of much more interest in most studies is the assessment of the qualities of good teachers. Reports on a number of these studies^{10,11,12} can be found in the USOE publication,

* Kiesling selected his California sample from projects already viewed as successful; thus, it is impossible to tell whether the same circumstances might not appear as frequently in unsuccessful programs, e.g., the result of looking at the expenditure level of the ten least successful State programs. Also, none of the large cities in California are included in his sample.

Do Teachers Make A Difference? Currently the findings are interpreted as showing that a teacher's degrees or units of extra study or even teaching experience is not as important as verbal ability.

Some skepticism appears appropriate on this point, especially with respect to teaching experience. For example, when Levin chose to price out the cost effectiveness of paying for verbal ability rather than experience, he used coefficients from Hanushek's analysis¹⁰ when coefficients from his own analyses would have shown the opposite conclusions. From the evidence currently at hand, teacher organizations will undoubtedly not rush to change salary formulas.

Undoubtedly, some teacher qualities are important in learning. For our purposes, suffice it to say that those variables most closely tied to level of per pupil expenditure, class size, or pupil/teacher ratio do not affect pupil outcomes. This finding is one of the best documented of all research findings. It is particularly significant because there is every likelihood that teachers will continue to press for reduced class size. If past experience is any guide, increases in fiscal resources will go mainly to increases in teacher salaries or reductions in class sizes.

In summary, the evidence from studies of school input does not indicate that increased educational expenditures per se will be effective in improving the disadvantaged student outcomes. Together with mechanisms to change school operating practices, they might well be effective, but neither the mechanism for change nor the specific changes required are self-evident at this time.

The results of our review confirm the lack of a significant and independent effect of school inputs, as predicted in the discussion of the model. Perhaps one of the reasons that such a condition exists is our fundamental lack of a cohesive body of information dealing with the

interaction of education and the society, the operative goals and allocation procedures in educational institutions, and the crucial elements in the teaching process. However, there is much that can be learned through analysis about the details of the educational process, with results that may be of limited use in improving the effectiveness of school expenditures in the short run, and might improve overall effectiveness in the long run.

RECOMMENDATIONS

A strong interdisciplinary attempt should be made to develop the best conceptual models and methodology for describing the educational process based upon what is already known. The work to date has generally been oriented to one discipline, methodologically unsophisticated (and often simply invalid), and well below the standard that could be reached even given the current lacks in the supporting findings of educational research.

Such a study would probably indicate the need for and potential benefits of new, specifically focused data collection efforts. However, it would also provide a referent against which to evaluate the biases expected in the use of the available data sources. We strongly recommend, whether new or existing data sources are used, that the potentialities of factor analysis or clustering techniques be used¹ to create useful macro variables within an area, e.g., school facilities, (2) to avoid problems of multicollinearity, and (3) to obviate the tendency to use stepwise regression or its variations.

Specific studies should look into the following:

- Tests for School-Sensitive Outcomes

Tests, measures, and indicators that are more sensitive to the effects of "good" formal education need to be created. Standardized tests are useful for some purposes, such as overall student comparisons, but the high interaction of their contents with student background factors blunt their usefulness in determining

the most effective alternative school inputs, policies, and practices. Moreover, if they were to become educational goals rather than merely correlates of the goals, as could happen with the increased emphasis on accountability, standardized tests are vulnerable to invalidation through "teaching to the tests."

Development of a next-generation of criterion-referenced tests to overcome some of their present limitations appears to be required.

- Non-Cognitive Process Measures

The interaction of school variables and students background variables appears to heavily involve other than cognitive aspects of student and teacher behavior. Current non-cognitive measures are many but inadequate, and better measures are needed, especially those that would spotlight the interaction of the sorting and teaching roles of the schools.

- Allocative Processes

Models of the allocative processes of educational institutions faced with multiple goals and changing priorities (often with very narrow institutional constraints) should be developed. Such models are crucial to the estimation of input parameters for subsets of outcome goals. This research might have the added benefit of indicating to schools the advantages of longer range educational programming.

- School-Parent Interactions

Information on the effectiveness of different school-based attempts to work with parents--e.g., the influencing background variables--should be drawn together. The work of Ira Gordon and Don Bushell, Follow Through project sponsors, highlights

the possibility of and benefits from school-based approaches to altering the child's home environment along certain limited dimensions that are particularly relevant to cognitive development.

It is not expected that the implementation of this set of recommendations will result in finding a dramatically larger role for school inputs in the overall educational outcomes for the disadvantaged. However, school inputs are an influence and these inputs will in total continue to cost multi-billions of dollars. Consequently, even small increases in our knowledge of how to use educational expenditures more efficiently could pay large dividends.

REFERENCES AND
OTHER SELECTED STUDIES

Note: "A" refers to individual multivariate studies, "B" to reviews of multivariate studies, and "C" to specific ESEA Title I studies.

1. Mayeske, G. W., et al., A Study of Our Nation's Schools, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, Washington, D.C., 1971. A
2. Levin, H. M., Report to the New York State Commission on Schools, 1971, 3 studies.
3. Stake, R. E., and J. L. Waldrop, Gain Score Errors in Performance Contracting, University of Illinois, Urbana-Champaign, Illinois.
4. Porter, O., Political and Administrative Factors Influencing the Allocation of Federal Aid: A Preliminary Summary of Interviews of State and District School Administrators, Syracuse University Educational Policy Research Center, Technical Memorandum Number 2, October 1968.
5. Wargo, M. J., et al., ESEA Title I: A Reanalysis and Synthesis of Evaluation Data from Fiscal Year 1965 Through 1970, American Institute for Research, Final Report CN OEC-0-71-4766, March 1972. C
6. Corder, Reginald, The Information Base for Reading, A Critical Review of the Information Base for Current Assumptions Regarding the Status of Instruction and Achievement in Reading in the United States, Educational Testing Service, Final Report Project Number 0-9031, Grant Number OEC-0-70-4792 (508), 1971.
7. Averch, H., S. J. Carroll and T. S. Donaldson, What Do We Know About Educational Effectiveness?, for President's Commission on School Finance, The Rand Corporation, Santa Monica, Calif., July 1971. B
(Note: Averch, Carroll and Donaldson provide a summary of the Unit of Analysis, Sample, Data, Variables, Analysis Procedures and Results for most of the individual multivariate studies.)

8. Thomas, J. A., Efficiency in Education: A Study of the Relationship Between Selected Inputs and Mean Test Scores in a Sample of Senior High Schools (unpublished Ph.D. dissertation, microf.), Stanford University Library, 1962. A
9. Burkhead, Jesse, G. Fox and W. Holland, Input and Output in Large City High Schools, Syracuse University Press, Syracuse, N.Y., 1967. A
10. Hanushek, Eric, "The Production of Education, Teacher Quality, and Efficiency" in Do Teachers Make a Difference?, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, OE-58042, Washington, D.C., 1970, pp. 79-99. A
11. Kiesling, H. J., Multivariate Analysis of Schools and Educational Policy, The Rand Corporation, P-4595, Santa Monica, Calif., March 1971. B
12. Michelson, Stephan, "The Association of Teacher Resourcefulness with Children's Characteristics," in Do Teachers Make a Difference?, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, OE-58042, Washington, D.C., 1970, pp. 120-168. A
13. Kiesling, H. J., Measuring a Local Government Service: A Study of School Districts in New York State, Review of Economics and Statistics. A
14. Parti, Michael and Irma Adelman, Optimal Allocation of Resources in Urban Education: An Econometric Approach, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, August 1971. A
15. Levin, H. M., "A New Model of School Effectiveness," in Do Teachers Make a Difference?, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, OE-58042, Washington, D.C., 1970, pp. 55-78. A
16. Mosteller, Frederick and Daniel Moynihan (eds), On Equality of Educational Opportunity, Vintage Book, New York, March 1972. B
17. Guthrie, J. W., et al., Schools and Inequality, The MIT Press, Cambridge, Mass., 1971. B
18. Cohn, Elchanan, "Economies of Scale in Iowa High School Operations," Journal of Human Resources, Vol. 4 (Fall 1968), pp. 422-434. A

19. Coleman, J.S., et al., "Pupil Achievement and Motivation," Chapter A
3, Equality of Educational Opportunity, U.S. Department of Health,
Education, and Welfare, U.S. Office of Education, OE-38001,
Washington, D.C., 1966.
20. Smith, M. S., Equality of Educational Opportunity: The Basic A
Findings Reconsidered, Center for Educational Policy Research,
Harvard Graduate School of Education, unpublished paper, Cambridge,
Mass., 1971.
21. Kiesling, H. J., A Study of the Cost and Quality of New York School A
Districts, Final Report, Project No. 8-0264, U.S. Department of
Health, Education, and Welfare, U.S. Office of Education,
Washington, D.C., 1970.
22. Ribich, T. I., Education and Poverty, The Brookings Institute, A
Washington, D.C., 1968.
23. Hanushek, E. A., The Education of Negroes and Whites, (unpublished A
Ph.D. dissertation, microf.) Massachusetts Institute of Technology
Libraries, Cambridge, Mass., 1968.
24. Kiesling, H. J., An Economic Analysis of Compensatory Education C
Programs in California, The Rand Corporation, Santa Monica, Calif.,
1971.

Other Individual Multivariate Studies

- Averch, Harvey, and H. J. Kiesling, The Relationship of School and A
Environment to Student Performance: Some Simultaneous Models for the
Project Talent High Schools, The Rand Corporation, Santa Monica,
Calif., forthcoming.
- Bensen, Charles, et al., "State and Local Fiscal Relationships in Public A
Education in California, Report of the Senate Fact Finding Committee
on Revenue and Taxation, Senate of the State of California, Sacramen-
to, March 1965.
- Fox, T. G., "School System Resource Use in Production of Interdependent A
Educational Outputs," The Joint National Meeting, American Astro-
nautical Society and Operations Research Society, paper presented
at Denver, Colorado, 1969.
- Katzman, M. T., "Distribution and Production in a Big City Elementary A
School System," Yale Economic Essays, Vol. 8, Spring 1968. pp. 201-
256.

- 17
- Kiesling, H. J., The Relationship of School Inputs to Public School Performance in New York State, The Rand Corporation, P-4211, Santa Monica, Calif., October 1969. A
- Mollenkopf, W. G., A Study of Secondary School Characteristics As Related to Test Scores, Research Bulletin RB-56-6, Educational Testing Service, Princeton, 1956. A
- Raymond, Richard, "Determinants of the Quality of Primary and Secondary Public Education in West Virginia," Journal of Human Resources, Vol. 3 Fall 1968, pp. 450-469. A

Other ESEA Title I Studies

- Wargo, M. J., P. Campeau and G. K. Tallmadge, Further Examination of Exemplary Programs for Educating Disadvantaged Children, American Institute for Research, Final Report CN OEC-0-70-5016, July 1971. C